Merging SenticNet and WordNet-Affect Emotion Lists for Sentiment Analysis

Soujanya Poria[1], Alexander Gelbukh[2], Erik Cambria[3], Peipei Yang[4], Amir Hussain[5], Tariq Durrani[6]

[1]*Computer Science and Engineering Department, Jadavpur University, India*
[2]*Center for Computing Research, National Polytechnic Institute, Mexico*
[3]*Temasek Laboratories, National University of Singapore, Singapore*
[4]*National Laboratory of Pattern Recognition, Chinese Academy of Sciences, China*
[5]*Dept. of Computing Science and Mathematics, University of Stirling, UK*
[6]*Dept. of Electronic and Electrical Engineering, University of Strathclyde, UK*

`soujanya.poria@gmail.com, gelbukh@gelbukh.com, cambria@nus.edu.sg`
`ppyang@nlpr.ia.ac.cn, ahu@cs.stir.ac.uk, durrani@strath.ac.uk`

Abstract—**SenticNet is currently one of the most comprehensive freely available semantic resources for opinion mining. However, it only provides numerical polarity scores, while more detailed sentiment-related information for its concepts is often desirable. Another important resource for opinion mining and sentiment analysis is WordNet-Affect, which in turn lacks quantitative information. We report a work on automatically merging these two resources by assigning emotion labels to more than 2700 concepts.**

Keywords—**Sentic computing, sentiment analysis, emotions.**

I. INTRODUCTION

Opinion mining is a rapidly developing area of natural language processing that has received a lot of attention from both research community and industry in recent years. The basic "feeling" about something can be described in terms of positive or negative polarity; thus the opinion detection task can be viewed as identifying the polarity of a piece of text.

For this purpose, one of the most comprehensive semantic resources currently available is SenticNet (Cambria *et al*. 2010). For about 5,700 concepts (single- or multi-word expressions), SenticNet gives a quantitative polarity score in the range between −1 (bad) to +1 (good), with neutral scores being around zero; e.g., *aggravation*: −0. 925, *accomplish goal*: +0.967, *December*: +0.111. This allows us to detect that a customer review is, say, 0.567 positive or 0.876 negative, according to the concepts appearing in the text.

However, more detailed information about a specific positive or negative emotion expressed in the text is often desirable. The main lexical resource employed currently to detect emotions in text is WordNet-Affect (Strapparava and Valitutti, 2004). For a relatively small number of words, it indicates whether a given word is related to one or more of six basic emotions, without information on how strong the emotion is: e.g., *angered* and *infuriated* have the same emotion label in WordNet-Affect.

There are much fewer concepts included currently in the WordNet-Affect lists than are contained in SenticNet. In this work, we automatically assigned specific emotion labels to a large number of SenticNet concepts that are not present in WordNet-Affect. This can be viewed both as adding new useful information to SenticNet entries and as expanding WordNet-Affect emotion lists with the SenticNet concepts. To the best of our knowledge, the resulting resource is the largest emotion lexicon currently available, and the first large quantitative emotion lexicon.

To assign emotion labels to SenticNet concepts, we train a classifier on the SenticNet concepts that are in WordNet-Affect and thus have their emotion labels already known. As the features for the classifier, we use a number of concept similarity measures, as well as features extracted for the occurrences of these concepts from an available emotion-related dataset. The paper is organized as follows: Section II gives a brief overview of the related work; Section III introduces the features used for classification; Section IV presents our classification procedure; Section V gives the evaluation results; finally, Section VI concludes the paper.

II. RELATED WORK

Our work lies in the intersection of two large and interrelated research fields: opinion mining and sentiment analysis.

A. *Opinion mining*

Opinion mining is a recent and explosively growing field of research. As the Web plays a more and more significant role in people's social lives, it contains more and more information concerning their opinions and sentiments. To effectively mine these opinions and detect the emotional content or positive/negative polarity of these opinions, there are three main approaches (Cambria and Hussain, 2012):

- keyword spotting: text is classified into categories according to the presence of unambiguous affect words;
- lexical affinity: arbitrary words are assigned a probabilistic affinity to a topic or emotion; and
- statistical methods: the automatically calculated valences of keywords, punctuation, and word co-occurrences frequencies calculated on a large training corpora are used.

All these approaches mainly rely on parts of text in which opinions and sentiments are explicitly expressed, such as polarity terms, e.g., *good, bad, nice, nasty, excellent, poor* and affect words, e.g., *happy, sad, calm, angry, interested, bored.* Accordingly, SenticNet was developed as a key source of polarity information by means of sentic activation (Cambria *et al.*, 2012a).

B. *Sentic Computing and SenticNet*

Currently available lexical resources for opinion polarity and affect recognition such as SentiWordNet (Esuli and Sebastiani, 2006) or WordNet-Affect are known to be rather noisy and limited. These resources mainly provide opinion polarity and affective information at syntactical level, leaving out polarity and affective information for common sense knowledge concepts such as *accomplish goal, bad feeling, celebrate special occasion, lose temper* or *be on cloud nine,* which are usually found in natural language text to express viewpoints and affect.

In contrast, recent approaches deal with concepts for mining the opinions. Sentic computing is a multidisciplinary approach to opinion mining and sentiment analysis at the crossroads between affective computing (Picard, 1997) and common sense computing, which combines computer and social sciences approaches to better recognize, interpret, and process opinions and sentiments found in the Web.

SenticNet is lexical resource built by clustering a vector space of affective common sense knowledge (Cambria *et al.*, 2011), which lists several thousand concepts along with their polarity. It was shown that, for example, for evaluation of medical patient data this lexicon based on concepts performed better than other available lexicons such as SentiWordNet or WordNet-Affect (Cambria *et al.*, 2012b).

C. *Affect and Emotions*

While emotions are not linguistic entities, the most convenient access that we have to them is through the language (Strapparava and Valitutti, 2004). Natural language texts not only contain informative contents, but also some attitudinal personal information including emotions, opinions, and attitudes. Recently, research activities related to emotions expressed in natural language texts and other media are gaining ground under the umbrella of subjectivity analysis and affect computing (Das, 2011).

The majority of subjectivity analysis methods related to emotions are based on textual keyword spotting using specially tailored lexical resources. A number of techniques for developing dictionaries of sentiment-related words have been proposed (Pang *et al.*, 2002; Wiebe and Mihalcea, 2006). The Affective lexicon (Strapparava and Valitutti, 2004) is one of the most important resources for detecting emotions in text, despite its small size.

Although a word may evoke different emotions in different contexts, an emotion lexicon will be a useful component for any sophisticated emotion detection algorithm (Mohammad and Turney, 2010). Accordingly, in this work we convert SenticNet into a large quantitative emotion lexicon.

TABLE I.
The data columns of ISEAR dataset used in this work

A. Background data related to the respondent: age; gender; religion; father's occupation; mother's occupation; country
B. General data related to the emotion felt in the situation described in the statement: intensity; timing; longevity
C. Physiological data: ergotropic arousals; trophotropic arousals, felt change in temperature
D. Expressive behavior data: movement, non-verbal activity; paralinguistic activity
E. Emotion felt in the situation described in the statement

III. FEATURES USED FOR CLASSIFICATION

The data items, or points, that we classified to assign the emotion labels (classes) to were concepts: single- or multi-word expressions present in the SenticNet vocabulary. When gathering information from the ISEAR corpus, we aggregated the information on multiple occurrences of the concept as a token in the text into one feature vector for the given concept as a type. For our classification, we used two kinds of features of the concepts: those based on the data columns directly provided in the ISEAR dataset, and those based on various similarity measures between concepts.

A. *ISEAR data-based features*

We used the 16 ISEAR data columns listed in Table I as independent features. We treated all the features we used, with the exception described below, as categorical features. For example, the *country* column has 16 different numerical codes, so we used 16 different features corresponding to each specific country. As the value, we used term frequency: if the concept occurs in the ISEAR dataset 3 times under country code 1 and 5 times under country code 2, then the corresponding part of the feature vector was (..., 3, 5, ...). The values expressing the degree or intensity of various parameters were, for simplicity, treated in the same way; there are around 3–4 discrete values per such data column in the ISEAR dataset. We did not use numeric data types for the values to avoid problems in aggregating (e.g., averaging) values for different occurrences of the same concept.

The only exception was the *age* column. We tested two different approaches: to treat all values (integer number of years) as different categorical values, or to group them in ranges. Grouping in ranges gave better results. According to (Wang *et al.*, 2011), a psychologically motivated grouping of ages is: 18–23 years, 23–28 years, and older than 28 years (all respondents of the ISEAR datasets were at least 18 years old), so we used these tree intervals as categorical features. This gave us about 100 categorical features, which were used as different dimensions of the feature vector.

B. *Features based on similarity measures*

Another kind of features is given by similarity measures between concepts. If we can measure somehow the similarity between pair of concepts, this gives us N distinct dimensions of the feature vector: the similarity between the given concept and each concept in the vocabulary; here N is the total number

of the concepts we considered, 2729 in our case. The intuition behind this is that if the distances from two data points in Euclidian space to a number of other points are similar, then it is probable that these two points are close to each other.

To define such similarity-based features for classification, we used the following 13 similarity measures: 10 lexical resources-based measures—SenticNet score-based similarity and nine WordNet-based similarity measures—and three co-occurrence-based measures—text distance-based similarity, point-wise mutual information, and emotional affinity, as described below. This gave us other $13N$ dimensions in the feature vectors.

The three co-occurrence-based measures (and in fact some of the WordNet similarity measures that incorporate corpus-based co-occurrence information) are highly correlated but still reflect different granularity levels of the text, so we decided to include all of them in the feature vectors.

1) SenticNet score-based similarity

The distance between two concepts a, b found in SenticNet is defined as $D_{SN}(a,b) = |p(a) - p(b)|$, where p(·) is the polarity specified for the concepts in SenticNet; the similarity is the inverse of the distance: $Sim_{SN}(a,b) = 1 / D_{SN}(a,b)$.

All concepts that we used were taken from SenticNet and thus have valid SentincNet scores.

2) WordNet distance-based similarity

The open-source package WordNet::Similarity (Pedersen *et al.*, 2004) with English WordNet 3.0 was used to measure the distance between tokens. This package provides nine similarity measures based on the analysis of the WordNet hierarchy, glosses, and other data present in WordNet. In this work, we used all the nine measures as independent sources of information, corresponding to its own N dimensions each in the feature vectors.

Not all concepts from SenticNet 1.0 are present in WordNet 3.0. Of those 2729 concepts that were found in the ISEAR database and with which we worked, 1436 were directly found in WordNet.

Those that were not found were examined manually and rephrased. For example, if a multi-word concept contained a main word, such as in *make mistake*, it was reduced to this word, *mistake* this case. After this, we could map 169 more concepts to WordNet, so in total 1605 concepts received meaningful WordNet similarity scores.

For the other 1124 concepts that we could not map to WordNet, the similarity between such a concept and any other concept was set to a random value. We did not set those values to zero or another fixed value because this would make all concepts not found in WordNet far from other concepts and very similar to each other, about 70% of coordinates ($9N$) in their feature vectors being exactly equal (we checked that doing so deteriorated the final results).

In contrast, setting them to random values better expresses the idea of unknown similarity: such vectors lie in the feature space rather far from all others and most importantly far from each other. Since all nine similarity scores are defined for specific senses and not for character strings, for a pair of concepts found in WordNet, the similarity was defined as the maximum similarity between all senses of the first one and all senses of the second one.

3) ISEAR text distance-based similarity

The positional information of the concept tokens in the ISEAR statements was used to measure the similarity between them. For this, we calculated the average minimum distance between the pairs of tokens that co-occurred in the statements of the ISEAR dataset.

Namely, if the tokens a and b occur in a statement S as words number $a_1, ..., a_n$ and $b_1, ..., b_n$, correspondingly, then the distance between them in this statement is defined as $D_{ISEAR}(a,b,S) = \min(|a_i - b_j|)$, and the distance over the entire ISEAR dataset is defined by averaging over individual statements S_k: $D_{ISEAR}(a,b) = \text{avg } D_{ISEAR}(a,b,S_k)$. Note that if the two words appear as a bigram, then $D_{ISEAR}(a,b,S) = 1$.

Now, the similarity is defined as the inverse of the distance: $Sim_{ISEAR}(a,b) = 1 / D_{ISEAR}(a,b)$. If the concepts do not co-occur in any statement, then we considered $Sim_{ISEAR}(a,b) = 0$.

4) Point-wise Mutual Information

The point-wise mutual information (PMI) between concepts measures the degree of co-occurrence between them within a sentence. For concepts a and b, it is defined as

$$Sim_{PMI} = \log \frac{p(a,b)}{p(a)p(b)}, \qquad (1)$$

where $p(a)$ is the probability for a sentence in the corpus to contain a, i.e., the number $n(a)$ of sentences where a occurs normalized by the total number of sentences in the corpus, and $p(a,b)$ is the probability for a sentence to contain both a and b, i.e., the number $n(a,b)$ of sentences that contain both a and b.

5) Emotional affinity

We define the emotional affinity between two concepts a and b in the same way as PMI but at the level of complete statements, i.e., $p(·)$ in (1) is defined as the corresponding number of statements instead of sentences, normalized by the total number of statements. Top ten emotional affinity pairs are given in Table II. The penultimate column in this table stands for the combination of the count of occurrences: $n(a,b) / n(a)n(b)$; see the previous subsection.

TABLE II.
Top 10 emotional affinity pairs

Concept pair		count	affinity
weekend	*December*	0.876	3.827
unemployed	*sad*	0.824	3.800
tired of	*headache*	0.765	3.768
happy	*December*	0.743	3.755
worry	*computer virus*	0.687	3.721
need relax	*annoyed*	0.665	3.707
mistake	*realize*	0.641	3.691
disgusting	*to person*	0.525	3.604
serious	*feel guilty*	0.430	3.518
birthday	*thought*	0.412	3.499

While PMI often reflects syntactic relatedness of the words—for example, it is high for a verb and its typical object, or for parts of a frequent multiword expression—emotional affinity incorporates a wider notion of relatedness within the same real-world situation, as well as synonymy and rephrasing. Due to our "one emotion per discourse" principle for the ISEAR dataset, the concepts with high emotional affinity tend to be related with the same emotion.

IV. CLASSIFICATION PROCEDURE

While several concepts appear in more than one WordNet-Affect list (for example, *harass* is listed under SADNESS and ANGER, *suspensive* under JOY and FEAR), the great majority most of them have only one emotion label. Predicting whether a word is ambiguous is outside the scope of the present paper, and we are not even sure if such ambiguity was not an error in the WordNet-Affect lists. Therefore, to simplifying things, we chose to assign only one emotion label to each concept.

In our evaluation, we consider a label to be assigned correctly if WordNet-Affect lists assign this label to the concept—even if it also assigns another label to it. There are too few cases of double labels present in the lists for this decision to significantly affect our results.

For classification, we used the SVM framework (Vapnik, 1995). Specifically, we used the `libsvm` library of the WEKA toolset (Hall *et al.*, 2009), which provides an implementation of a multiclass SVM. As a result, we obtained one emotion label for each concept.

A set of standard pre-processing techniques such as tokenizing and lemmatizing were used in the process as described below. For this, we used the tools provided by Rapidminer's text plug-in, except for lemmatizing, for which the WordNet lemmatizer (Miller, 1995) was used (a lemmatizer differs from a stemmer in that it provides a complete form: for *feet*, it provides *foot*).

For each SenticNet concept, we identified all its occurrences in the text of ISEAR statements. For multi-word SenticNet concepts, such as *after summer*, *to person*, etc., we allowed any number of stop-words to appear the position of the space, so that in the SenticNet vocabulary matched, e.g., *to a person* or *to the person* in the text.

Lemmatizing both the SenticNet concepts and the ISEAR text before matching would generate some number of false matches, while not using any lemmatization would result in many words in the text to fail to match with the concepts because they appear in the text in a different form, e.g., *made mistake* in the text vs. *make mistake* in the vocabulary. To minimize both undesired effects, for each concept in the SenticNet vocabulary and for each ISEAR statement, we first tried to find the concept in the statement; if zero occurrences of the concept were found in this particular statement, then we lemmatized both the concept and all words in the statement and repeated the search.

A total of 2729 SenticNet concepts were found at least once in the ISEAR dataset. Only these concepts participated in further processing and were finally assigned the emotion labels. For each occurrence, the corresponding data fields were extracted from the ISEAR dataset, and the data for multiple occurrences of the same concept in the whole corpus were aggregated in a feature vector for that concept. This gave us a dataset with a total of 2729 feature vectors.

All those 2729 concepts were assigned the emotion labels. However, as training and test data, we used the intersection between the sets of concepts found in the WordNet-Affect lists (for which we had the gold standard emotion labels) and those 2729 SenticNet concepts found in the ISEAR texts (for which we had ISEAR-specific features); this intersection consisted of 1202 concepts.

V. EVALUATION AND EXPERIMENTAL RESULTS

For evaluation, a standard tenfold cross-validation procedure was carried out. We have conducted several experiments to determine the impact of different features. In particular, we observed that the use of the similarity measures improved our results in comparison with using only ISEAR data-based features. We also studied the impact of various groups of ISEAR data-based features. The most instructive results are shown in Table III.

In Table III, S stands for the use of the features based on SenticNet similarity, W for WordNet-based similarity measures, L for the ISEAR text-based similarity, and A to D for the groups of ISEAR data shown in Table I.

It can be observed that the use of SenticNet-based similarity had a positive impact. The use of all available features gave best results both when only similarity-based features were used and when ISEAR data-based features were taken into account. With all features, the accuracy of 85.12% was achieved. Analysis of the correlation of the ISEAR's general and physiological variables with emotions has previously shown interesting insights: e.g., low intensity for emotion classes of SHAME and GUILT and high for JOY, FEAR and SADNESS (Das and Bandyopadhyay, 2011).

TABLE III
Precision with different feature combinations

S	W	T	A	B	C	D	Precision %
+							29.07
	+						50.59
+	+						57.77
		+					*27.87*
+		+					28.21
	+	+					43.09
+	+	+					**59.27**
+	+	+	+				73.23
+	+	+		+			65.12
+	+	+			+		60.22
+	+	+				+	58.00
+	+	+	+	+			61.12
+	+	+	+		+		60.24
+	+	+	+			+	58.45
+	+	+		+	+		*56.34*
+	+	+		+		+	66.12
+	+	+	+	+	+	+	**85.12**

As can be now observed from Table III, the expressive behavior contributes in the detection of emotion less than other groups of variables. We also evaluated how well our algorithm identifies positive or negative emotions irrespective of a specific emotion label. Considering JOY and SURPRISE to be both positive emotions—even if the gold standard indicates one of them and our algorithm assigned the other—and the rest negative, we observed 96.11% agreement with the gold standard—the WordNet-Affect emotion lists.

The following two examples illustrate our procedure. In the text ISEAR statement "*When I made the same mistake that I had accused someone else of, and this was obvious to a third pereson*" we manually corrected the typo in the word *person*, after which the following SenticNet concepts were identified in it: *make mistake*, *obvious*, *to person*, to which our classifier assigned the labels SADNESS, JOY, SURPRISE, correspondingly; note that labels are assigned to concepts as types, not to the specific occurrences of these concepts in this text.

Another example: in the statement "*When friends try to put me down or hurt me*"; the following concepts were found, to which the following labels were finally assigned: *friend* → JOY, *put down* → SADNESS, *hurt* → FEAR.

We identified two main sources of errors. One consists in not properly taking into account co-occurrence of related words in a statement when they appear in different sentences; the algorithm failed to transfer information from one word to the other. The other source of errors is word sense ambiguity.

VI. CONCLUSIONS AND FUTURE WORK

We have extended the emotion labels indicated in the SemEval 2007 WordNet-Affect emotion lists to 2729 concepts from the SenticNet resource. With this, to the best of our knowledge, we created the largest resource marked with emotion labels, as well as the first quantitatively marked emotion lexicon. We also studied the impact of different psychological features on emotion classification.

The work opens a number of directions for future work. The most obvious directions include using other types of corpora as a source of features to improve the coverage and thus label a greater number of concepts, as well as the use of more elaborated classification techniques in combination with syntactic and psychological clues to improve accuracy.

We will also apply the resource we built to real-life opinion mining tasks, such as mining the patient data in the way in which the efficiency of SenticNet has already been demonstrated.

REFERENCES

[1] Cambria, E., Speer, R., Havasi, C., Hussain, A. 2010. SenticNet: A publicly available semantic resource for opinion mining. In Proceedings of AAAI CSK, pp. 14-18, Arlington.

[2] Cambria, E., Hussain, A. 2012. Sentic Computing: Techniques, Tools, and Applications. Dordrecht: Springer.

[3] Cambria, E., Olsher, D., Kwok, K. 2012a. Sentic activation: A two-level affective common sense reasoning framework. In: Proceedings of AAAI, pp. 186-192, Toronto.

[4] Cambria, E., Mazzocco, T., Hussain, A., Eckl, C. 2011. Sentic medoids: Organizing affective common sense knowledge in a multi-dimensional vector space. LNCS, vol. 6677, pp. 601-610. Springer-Verlag, Berlin Heidelberg

[5] Cambria, E., Benson, T., Eckl, C., Hussain, A. 2012b. Sentic PROMs: Application of sentic computing to the development of a novel unified framework for measuring health-care quality. Expert Systems with Applications 39(12), pp. 10533-10543, Elsevier

[6] Das D., 2011. Studies on Emotion Analysis at Word and Sentence Level. Lambert Academic Publishing, pp. 96.

[7] Das, D., Bandyopadhyay, S. 2011. Analyzing Emotional Statements – Roles of General and Physiological Variables. In: SAAIP Workshop, 5th IJCNLP, pp. 59-67, Chiang Mai.

[8] Ekman, P. 1993. Facial expression and emotion. American Psychologist, 48(4): 384-392.

[9] Esuli, A., and Sebastiani, F. 2006. SentiWordnet: A publicly available lexical resource for opinion mining. In: Proceedings of LREC.

[10] Gale, W., Church, K., Yarowsky, D. 1992. One Sense per Discourse. In Proceedings of the 4th DARPA Speech and Natural Language Workshop, pp. 233-237.

[11] Hall, M., Frank, E., Holmes. G., Pfahringer, B., Reutemann, P., Witten, I.H. 2009. The WEKA Data Mining Software: An Update; SIGKDD Explorations 11(1).

[12] Miller, G.A. 1995. WordNet: A lexical database for English. Communications of the ACM, 38(11): 39-41.

[13] Mohammad, S., Turney, P.D. 2010. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. Proceedings of the NAACL-HLT, Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, pp. 26-34.

[14] Ortony, A., Turner, T. J. 1990. What's basic about basic emotions? Psychological Review 97, 315-331.

[15] Pang, B., Lillian, L., Shivakumar, V. 2002. Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of EMNLP, pp. 79-86.

[16] Pedersen, T., Patwardhan, S., Michelizzi, J. 2004. WordNet::Similarity — Measuring the Relatedness of Concepts. In: Proceedings of HLT-NAACL: Demonstration Papers, Association for Computational Linguistics, pp. 38-41.

[17] Picard, R. 1997. Affective Computing. The MIT Press

[18] Scherer K.R. 2005. What are emotions? And how can they be measured? Social Science Information, 44(4): 693-727.

[19] Strapparava, C., and Valitutti, A. 2004. WordNet-Affect: an affective extension of WordNet. In: Proceedings of LREC.

[20] Vapnik, V.N. 1995. The Nature of Statistical Learning Theory, Springer-Verlag, pp. 187.

[21] Wang, W., Kathol, A., Bratt, H. 2011. Automatic Detection of Speaker Attributes Based on Utterance Text. In: Proceedings of INTERSPEECH, pp. 2361-2364.

[22] Wiebe, J., Mihalcea, R. 2006. Word sense and subjectivity. In: Proceedings of COLING/ACL, pp. 1065-1072, Sydney.